

The coverage factor in a Flatten–Gaussian distribution

J Blázquez¹, A García-Berrocal^{2,3}, C Montalvo² and M Balbás²

¹ Nuclear Fission Division, CIEMAT, Avenida Complutense 22, 28040 Madrid, Spain

² AMERPREM Research Group, School of Mines, Technical University of Madrid (UPM),
Ríos Rosas 21, 28003 Madrid, Spain

E-mail: agustin.garciaberrocal@upm.es

Abstract

According to the *Guide to the Expression of Uncertainty in Measurement*, a coverage factor that produces an expanded uncertainty having an approximate level of confidence is recommended in the final expression of the measurement result. The numerical value for the coverage factor depends on the probability density function. When a Type A evaluation following a Gaussian (normal) distribution is combined with a Type B one following a rectangular distribution, the resulting probability density function is neither Gaussian nor rectangular, but similar to both, which we term a Flatten–Gaussian probability density function. The variance of such a function is the combined variance in the usual way, but the coverage factor must be calculated.

1. Introduction

Following the recommendation of the *Guide to the Expression of Uncertainty in Measurement (Guide)* [1] a specific level of confidence must be associated with the interval defined by the expanded uncertainty. This requires certain assumptions regarding the probability distribution characterized by the measurement result and its combined standard uncertainty. The level of confidence that may be attributed to this interval can be known only to the extent to which such assumptions may be justified.

The procedure given in the *Guide* [1, section G.4] recommends an approximated method in order to solve this problem. Nevertheless, in those cases in which the probability distribution characterized by the measurement result and its combined standard uncertainty may be approximated by a t -distribution whose effective degrees of freedom are considerably high, it is convenient to suppose that a $k = 2$ coverage factor produces an interval having a level of confidence of approximately 95% and that a $k = 3$ one produces an interval having a level of confidence of approximately 99%.

In most cases, a measurand is not measured directly, but is determined from other input quantities through a functional relationship. In this paper we consider the measurand as an

addition of two input quantities, one of them being a probability distribution based on a series of observations and the other one an *a priori* distribution.

The *Guide* [1] establishes the Type A denomination for the evaluation of the standard uncertainty component based on frequency distributions while Type B evaluation is founded on the *a priori* distribution, and the *Guide* recommends the estimated standard deviation associated with the measurement result (combined standard uncertainty) to be obtained by means of the approximation named as the *law of propagation of uncertainty*. Then the coverage factor used to obtain the expanded uncertainty must be calculated from the resulting probability distribution for the measurand.

In this work, we assume that the Type A evaluation is founded on a Gaussian (normal) distribution and the Type B evaluation is a symmetrical rectangular distribution. At a first approach, the coverage factor may be estimated as a linear combination of both Gaussian and rectangular coverage factors, weighted with their respective standard deviations. This approach can be tested by building a Monte Carlo combination of Gaussian and rectangular probability distribution functions. The resulting probability density function (pdf), here called Flatten–Gaussian, obviously depends on two parameters, which are a combination of the original variances. For instance, the variance of the Flatten–Gaussian pdf is the sum of the original variances.

Extreme cases arise when one of the two original variances is dominant, then the Flatten–Gaussian pdf is very much like the dominant distribution. But when both original variances are the same order of magnitude, the Flatten–Gaussian pdf inherits features of both distributions, so the resulting coverage factor must be researched.

The coverage factor could also be calculated using the polynomial chaos approach [2, 3]. This is a very general approach, valid for many distributions; furthermore, they can be generalized to multivariate cases. In spite of the fact that the polynomial chaos method is an established approach [4], it is not yet widely used by practitioners; so we have chosen the maximum entropy approach for calculating the coverage factor.

An exact analytical form of the Flatten–Gaussian pdf (noted by $p_{FG}(\cdot)$) is hard to find because the rectangular distribution is not analytical at the borders; although an approximated analytical form can be built by using the maximum entropy principle constrained by higher order central moments [5]. Up to order four, the resulting analytical form is $p_{FG}(Y) = A \exp(-aY^2 - bY^4)$. Once the parameters a and b are found, a coverage factor can be calculated for such a distribution.

In this paper, the above analytical form is fitted to a Monte Carlo generated Flatten–Gaussian pdf. Parameters a and b are determined from the fit, and the coverage factor is calculated. The sign of the first parameter reflects when the four central moments approach is sufficient. The calculation of the two parameters, regarded as Lagrange multipliers, is not straightforward; here, easier procedures are designed for practical purposes.

2. The Flatten–Gaussian distribution

Let X_1 be a quantity with uncertainty following a Gaussian pdf with x_1 as the expected value and $u(x_1)$ as the standard deviation, that is Type A standard uncertainty, and let X_2 be a rectangular pdf with x_2 as the expected value and $u(x_1)$ as the standard deviation, that is, Type B standard uncertainty. Finally, letting $Y = f(X_1, X_2)$ be the measurand, what is the pdf $p_{FG}(Y)$ for estimating the uncertainty of the expected output value when $Y = X_1 + X_2$?

Being σ_Y^2 the combined variance of the estimate y , it is well established that $\sigma_Y^2 = \sigma_g^2 + \sigma_r^2$, where

$$\sigma_g^2 = \left(\frac{\partial f}{\partial X_1} \right)^2 u^2(x_1), \quad \sigma_r^2 = \left(\frac{\partial f}{\partial X_2} \right)^2 u^2(x_2).$$

In the extreme case that $\sigma_g^2 \gg \sigma_r^2$ the pdf for Y can be regarded as Gaussian, in the other extreme, $\sigma_g^2 \ll \sigma_r^2$, as rectangular. But when the variances are the same order of magnitude, the probability distribution is intermediate, so here it is called ‘Flatten–Gaussian’ and its analytical form must be established in order to calculate the coverage factor.

The problem is already solved using two approaches: central moments [5] and polynomial chaos [6]. For an analytical solution, these approaches are needed because the rectangular distribution is not analytical at its borders. We

shall follow the central moments approach, keeping in mind that the odd moments must be zero due to the symmetry of the Flatten–Gaussian pdf; hence, only the second and fourth central moments are considered.

The moment generating function for the Flatten–Gaussian distribution must be

$$M_Y(t) = \exp\left(\frac{\sigma_g^2 t^2}{2}\right) \frac{\sinh(a_r t)}{a_r t}, \quad a_r = \sqrt{3}\sigma_r,$$

where the first factor corresponds to the Gaussian distribution and the second to a rectangular distribution of half-width a_r . Developing $M_Y(t)$ in a Taylor series up to the fourth order, the second and fourth moments are obtained:

$$\mu_2 = \sigma_g^2 + \sigma_r^2; \quad \mu_4 = 3\sigma_g^4 + 6\sigma_g^2\sigma_r^2 + \frac{9}{5}\sigma_r^4.$$

Observe that μ_2 is the sum of the variances, as expected; but μ_4 contains a crossing term which is the dominant one in the case of both variances being alike. Besides, $1.8 \leq \mu_4/\mu_2^2 \leq 3$. When both variances are equal, the ratio of the central moments is 2.7, quite close to 3, the Gaussian reference; so, in that case, the Flatten–Gaussian pdf approaches a single Gaussian with increased variance μ_2 . That has been done traditionally; although, when this ratio becomes lower because $\sigma_g < \sigma_r$, the traditional approach must be corrected by using the proper probability distribution.

The analytical form of the Flatten–Gaussian pdf is obtained applying the principle of maximum entropy constrained with the four moments [7]. The result is

$$p_{FG}(Y) = A \exp(-aY^2 - bY^4),$$

where A is the normalization factor to unit area and the coefficients a and b are obtained as Lagrange multipliers making the entropy maximum [8].

Some restrictions for the coefficients are required: for convergence reasons, $b > 0$. Nevertheless, the coefficient a , regarded as a Lagrange multiplier, could be negative. In this case, the Flatten–Gaussian distribution would not be unimodal. For avoiding this case, $a > 0$ is always required. The case when $a < 0$ corresponds to $\sigma_g \ll \sigma_r$; hence, it is preferable to approach the pdf by a rectangular instead of a Flatten–Gaussian distribution.

The normalization factor A is calculated from

$$\int_{-\infty}^{\infty} p_{FG}(Y) dY = 1.$$

Denoting $\alpha = ab^{-1/2}$, it is found that

$$\begin{aligned} & \int_{-\infty}^{\infty} \exp(-aY^2 - bY^4) dY \\ &= 2b^{-1/4} \int_0^{\infty} \exp(-\alpha u^2 - u^4) du \\ &= 2b^{-1/4} \left[\frac{1}{4} \Gamma\left(\frac{1}{4}\right) {}_1F_1\left(\frac{1}{4}; \frac{1}{2}; \frac{\alpha^2}{4}\right) \right. \\ & \quad \left. - \alpha \frac{1}{4} \Gamma\left(\frac{3}{4}\right) {}_1F_1\left(\frac{3}{4}; \frac{3}{2}; \frac{\alpha^2}{4}\right) \right], \end{aligned}$$

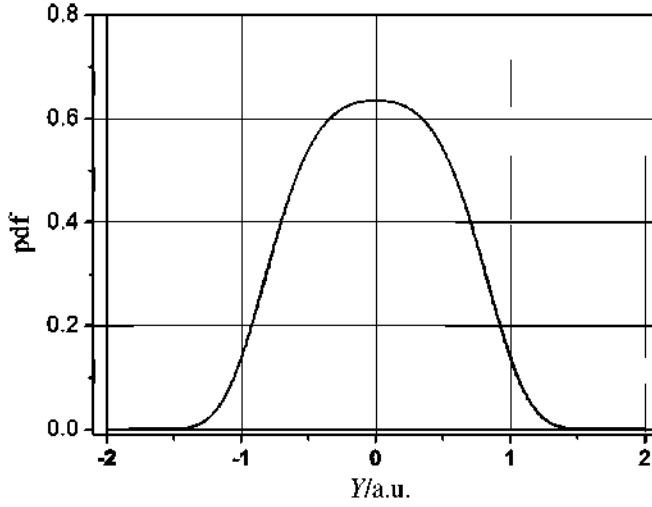


Figure 1. The shape of a Flatten-Gaussian pdf for the parameters $a = 0.35$ and $b = 1.17$.

with ${}_1F_1(n; m; x)$ the first order confluent hypergeometric function [9], defined by

$${}_1F_1(n; m; x) = 1 + \frac{n}{m}x + \frac{n(n+1)}{m(m+1)} \frac{x^2}{2!} + \frac{n(n+1)(n+2)}{m(m+1)(m+2)} \frac{x^3}{3!} + \dots,$$

and furthermore:

$$\Gamma\left(\frac{1}{4}\right) \approx 3.62561; \quad \Gamma\left(\frac{3}{4}\right) \approx 1.22542.$$

As a result, making the change in variables: $\lambda = b^{-1/4}$, $s = (2a)^{-1/2}$, the complete form of the Flatten-Gaussian pdf is

$$p_{FG}(Y) = \frac{\left(\frac{2}{\lambda}\right) \exp\left(-\frac{Y^2}{2s^2} - \frac{Y^4}{\lambda^4}\right)}{\Gamma\left(\frac{1}{4}\right) {}_1F_1\left(\frac{1}{4}; \frac{1}{2}; \frac{\lambda^4}{16s^4}\right) - \frac{\lambda^2}{2s^2} \Gamma\left(\frac{3}{4}\right) {}_1F_1\left(\frac{3}{4}; \frac{3}{2}; \frac{\lambda^4}{16s^4}\right)}.$$

In figure 1, $p_{FG}(Y)$ is plotted for $\lambda = 0.961$, $s = 1.19$, the ratio $(\lambda/2s)^4 = 0.026$ meaning that the rectangular form is the dominant, although the pure Gaussian distribution cannot be neglected.

Even central moments of a Flatten-Gaussian distribution can be calculated by using

$$\begin{aligned} \int_0^\infty u^{2k} \exp(-\alpha u^2 - u^4) du \\ = \frac{1}{4} \left[\Gamma\left(\frac{2k+1}{4}\right) {}_1F_1\left(\frac{2k+1}{4}; \frac{1}{2}; \frac{\alpha^2}{4}\right) - \alpha \Gamma\left(\frac{2k+3}{4}\right) {}_1F_1\left(\frac{2k+3}{4}; \frac{3}{2}; \frac{\alpha^2}{4}\right) \right]. \end{aligned}$$

In order to obtain a coverage factor for the Flatten-Gaussian pdf a pair of constants a and b are required, although the calculation of the constants is not a straightforward process. The constants can be found as Lagrange multipliers in the constrained entropy minimization process. They can also be found by equating the second and fourth moments, obtained

Table 1. Coverage factors for several distributions: k_v is the averaged coverage factor and k_e is the estimate of coverage factor using the Monte Carlo method.

σ_g	σ_r	k_v	k_e
1	0	1.96	1.99
1	1.15	1.79	1.86
1	1.73	1.73	1.84
1	2.88	1.71	1.75
0.05	0.57	1.66	1.62

from the raw measurements, with the moments using the probability distribution. In both cases, a non-linear system of equations must be solved.

In the case of $(\lambda/2s)^4 \ll 1$, the order fourth term in the argument of the exponential is the dominant, and the distribution function tends to a rectangular distribution with a variance equal to $\lambda^2/3$:

$$p_{FG}(Y) \approx \frac{\left(\frac{2}{\lambda}\right) \exp\left(-\frac{Y^4}{\lambda^4}\right)}{\Gamma(1/4)}.$$

This consideration stands for choosing initial values required in solving a non-linear system of equations. For our case $\sigma_r > \sigma_g$ the approximate solutions are

$$\lambda = \sqrt{3\mu_2}; \quad s^2 = \sqrt{\frac{\mu_4}{6(\sigma_g/\sigma_r)^2 + (9/5)}}.$$

3. Empirical probability density functions

By generating random numbers uniformly distributed in $(0, 1)$ empirical Flatten-Gaussian functions can be built [10]. Let ξ_1 and ξ_2 be such numbers; then the random variable Y , defined by

$$Y = \frac{C_1}{1.855} \ln\left(\frac{1 - \xi_1}{\xi_1}\right) + C_2(2\xi_2 - 1),$$

follows a Flatten-Gaussian distribution. The constants C_1 and C_2 are fitted for having σ_g and σ_r respectively. The empirical coverage factor k_e for 95% of probability is calculated. Besides, an averaged coverage factor k_v is defined as

$$k_v = \frac{1.64\sigma_r + 1.96\sigma_g}{\sigma_r + \sigma_g}.$$

A comparison between both coverage factors is shown in table 1 for several Flatten-Gaussian distributions. As a consequence, calculating the averaged coverage factor is satisfactory for practical purposes with a 2.5% discrepancy.

Empirical pdf is an alternative way of estimating the constants λ and s^2 for the theoretical Flatten-Gaussian distribution, without finding the Lagrange multipliers. Constants are obtained by solving the least squared problem:

$$\ln[p_{FG}(Y)] = \ln A - \frac{Y^2}{2s^2} - \frac{Y^4}{\lambda^4}.$$

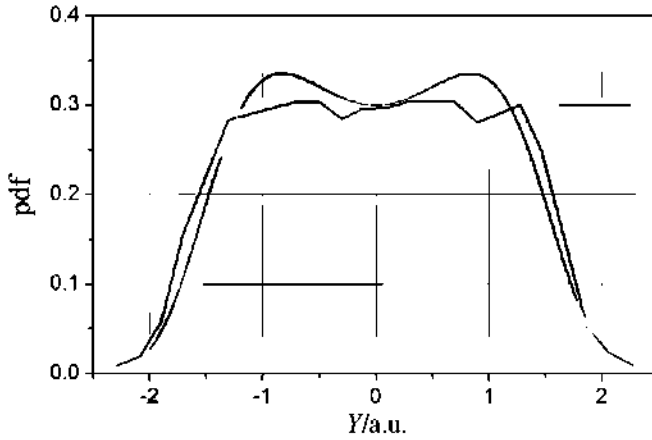


Figure 2. Bimodal Flatten–Gaussian pdf for $\sigma_r > 4\sigma_g$; the black line results from Monte Carlo simulation and the grey line is the best least squared fit.

The fit loses its sense when it yields a negative value for the coefficient of Y^2 , because s^2 must always be positive, which happens when $\sigma_r > 4\sigma_g$; the rectangular distribution is much broader than the Gaussian, so the resulting Flatten–Gaussian is not unimodal for a better fit. This can be corrected adding the sixth central order moment to the maximum entropy procedure, which in turn gives a new parameter to the fit. The algebra becomes rather cumbersome, with no practical interest. In such conditions, to calculate the coverage factor as if the Flatten–Gaussian pdf were a pure rectangular distribution is the best choice. See figure 2.

4. Conclusions

The Flatten–Gaussian probability density function appears as a consequence of combining Type A and Type B standard uncertainties. The 95% coverage factor for such a distribution has been researched.

The extreme cases, when Type A or Type B is dominant, are treated neglecting the coverage factor of the non-dominant type, but using the second moment as the variance. Intermediate cases when $1 < \sigma_r/\sigma_g < 4$ require an analytical form for the Flatten–Gaussian distribution in order to calculate the proper coverage factor.

On finding the maximum entropy constrained with the second and fourth central moments the resulting analytical form is

$$p_{FG}(Y) = \frac{\left(\frac{2}{\lambda}\right) \exp\left(-\frac{Y^2}{2s^2} - \frac{Y^4}{\lambda^4}\right)}{\Gamma\left(\frac{1}{4}\right) {}_1F_1\left(\frac{1}{4}; \frac{1}{2}; \frac{\lambda^4}{16s^4}\right) - \frac{\lambda^2}{2s^2} \Gamma\left(\frac{3}{4}\right) {}_1F_1\left(\frac{3}{4}; \frac{3}{2}; \frac{\lambda^4}{16s^4}\right)}.$$

There are three procedures for estimating the constant λ and s^2 .

1. Lagrange multipliers in finding the maximum entropy.
2. Calculating the moments with the above analytical form and equating to the empirical moments.
3. By Monte Carlo simulation and solving a least squared fit.

The first and second procedures require initial values for the parameters. The Monte Carlo is the easiest procedure by far.

The averaged coverage factor is compared with the exact one. Discrepancies are within 2.5%; hence, its use is recommended for practical purposes, keeping in mind that the evaluation of Type B uncertainties with a rectangular distribution is not exact.

Acknowledgments

This work has been performed within the research programme of Cátedra CLH de Metrología de los Hidrocarburos. This programme was developed with the scientific collaboration between the Technical University of Madrid (UPM) and the Compañía Logística de Hidrocarburos CLH. The authors would like to thank Dr P Vilarroig for a careful review of the manuscript.

References

- [1] BIPM, IEC, IFCC, ISO, IUPAC and OIML 1995 *Guide to the Expression of Uncertainty in Measurement* (Geneva: International Organization for Standardization)
- [2] Wiener N 1938 The homogeneous chaos *Am. J. Math.* **60** 897–936
- [3] D'Antona G, Monti A, Ponci F and Rocca L 2007 Maximum entropy multivariate analysis of uncertain dynamical systems based on the Wiener–Askey polynomial chaos *IEEE Trans. Instrum. Meas.* **56** 689–95
- [4] D'Antona G, Monti A, Ponci F and Rocca L 2006 Maximum entropy analytical solution for stochastic differential equations based on the Wiener–Askey polynomial chaos *IEEE Proc. AMUEM (Sardagna, Trento, Italy)* pp 62–6
- [5] D'Antona G 2004 Expanded uncertainty and coverage factor computation by higher order moments analysis *IEEE Proc. IMTC (Como, Italy)* pp 234–8
- [6] Lovett T and Monti A 2005 A polynomial chaos approach to measurement uncertainty *IEEE Proc. AMUEM (Niagara Falls, Ontario, Canada)* pp 33–8
- [7] Einbu J M 1977 On the existence of a class of maximum-entropy probability density functions *IEEE Trans. Inform. Theory* **23** 772–5
- [8] Marlow W H 1995 *Mathematics for Operation Research* (New York: Dover)
- [9] Nikiforov A F, Suslov S K and Uvarov V B 1991 *Classical Orthogonal Polynomials of a Discrete Variable* (Berlin: Springer)
- [10] Fröberg C E 1979 *Introduction to Numerical Analysis* 2nd edn (Reading, MA: Addison Wesley)